



INFORMATION EXTRACTION TOOL (IET)

**John Baumler ([baumlerj@rl.
af.mil](mailto:baumlerj@rl.af.mil))
AFRL/IFEB
315-330-3979 (DSN587-3979)
27 Nov 2000**



Background

- DODIIS AMHS provides message archiving, retrospective search and message generation/review capabilities for organizational messages (DMS Will Lack This Capability)
- 497IOG is the DODIIS Executive Agent for DODIIS AMHS
 - 26 legacy DODIIS AMHS sites
 - 2100/4100 hardware requires recap over next two years per COMPAQ
 - 5 DAWS Message Front End (DMFE) sites maintained by AFRL



Background (continued)

- Dec 98 - 497IG/IND requested AFRL develop a low cost automated message handling capability for the SCI DMS-era architecture
 - Add BROADSWORD as user interface
 - Provide migration path to DMS user agent (Outlook)
 - Support DODIIS Target Operating Systems (Solaris/Windows NT)
 - Selection Based on DMS Implementation, Functional Equivalence and Life Cycle Cost Analysis
 - Functional Requirements Comparison Available on IET Web Page



Information Extraction Tool (IET)

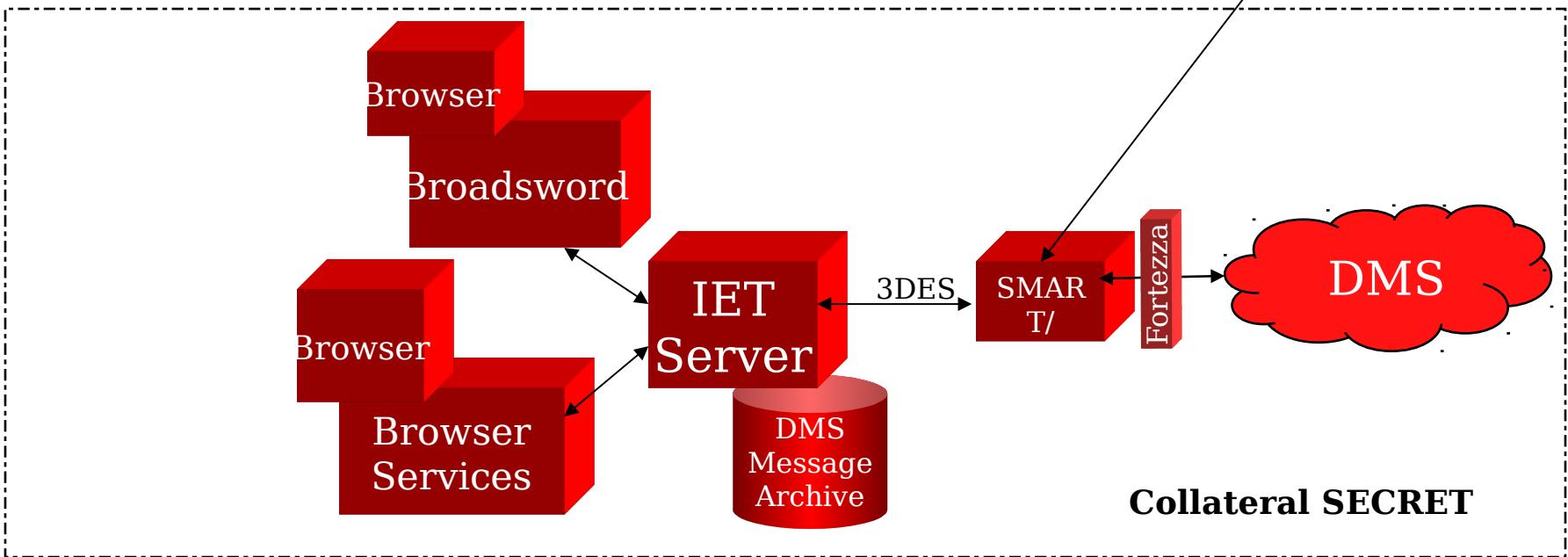
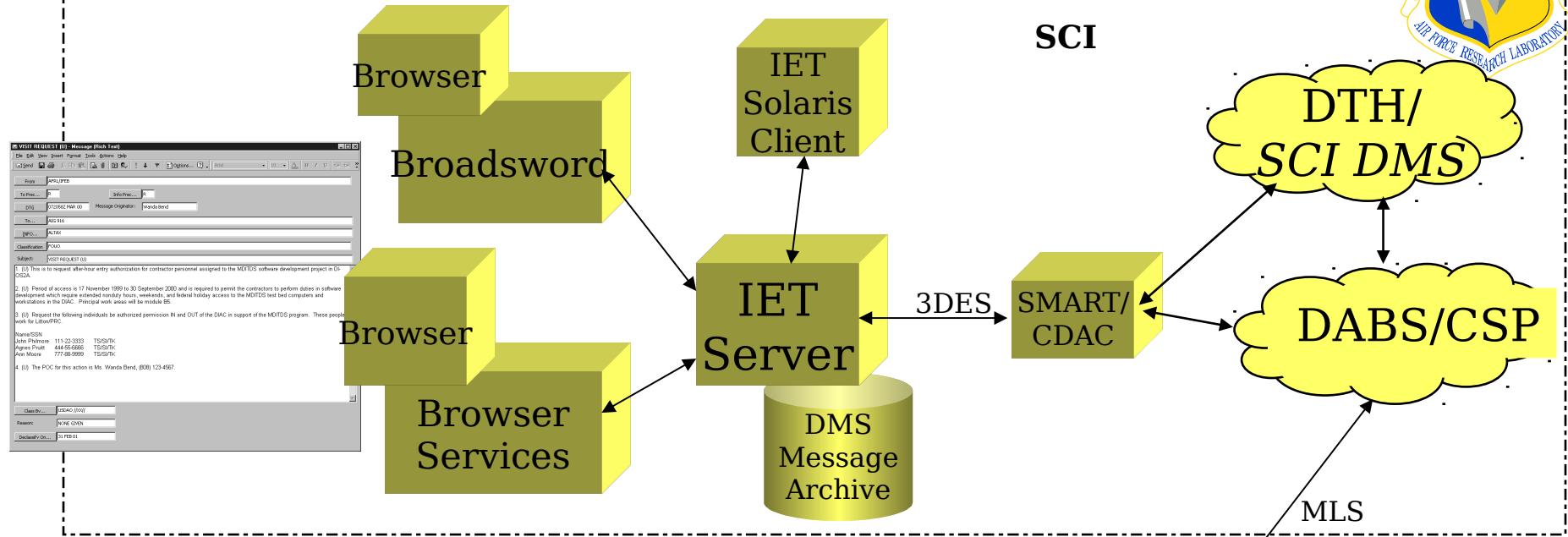
- Value added profiling and search capabilities for the intelligence analyst; supplements DMS architecture
- Provides analyst with access to archived messages (and enclosures-v1.1)
- Provides transitional message generation functionality
- Integrates into DMS-era architecture by interfacing with CSP/CDAC
- Maximizes use of existing accredited interfaces
- Supports local and remote user base
- Serves as platform for future integration of advanced information extraction technologies



Information Extraction Tool (IET)

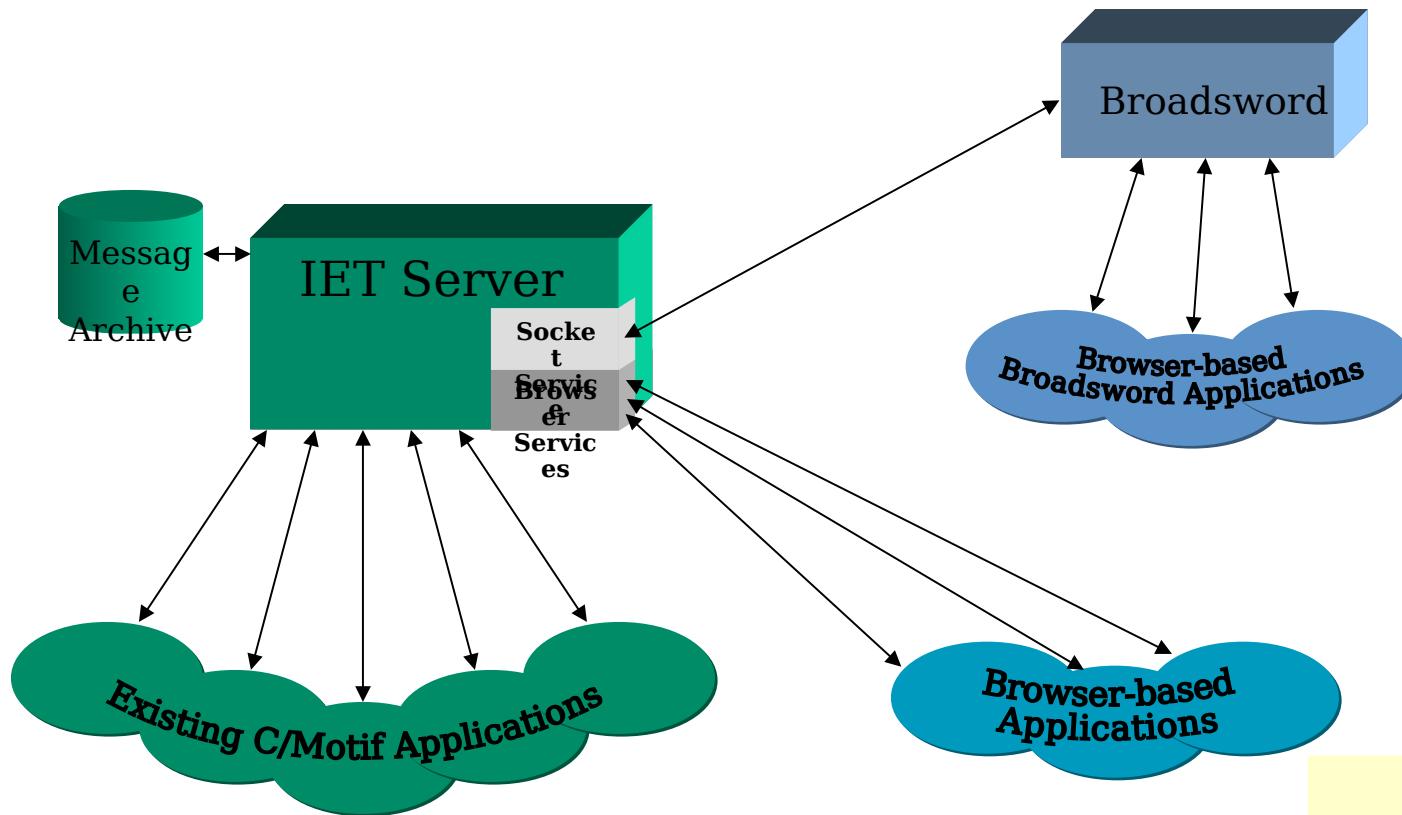
- DMFE provides IET core services
 - Sun Solaris server environment, Sybase DBMS
- Adds Browser-based interface
- BROADSWORD plug-in available for profiling/retrospective search

IET as Part of DMS Architecture





IET Client Environment



- C/Motif applications run on Solaris clients

- Broadsword message access through socket services

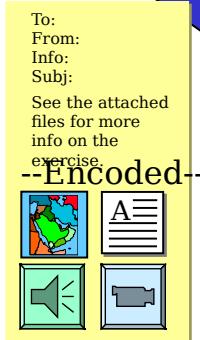
- Browser-based message access, display and generation



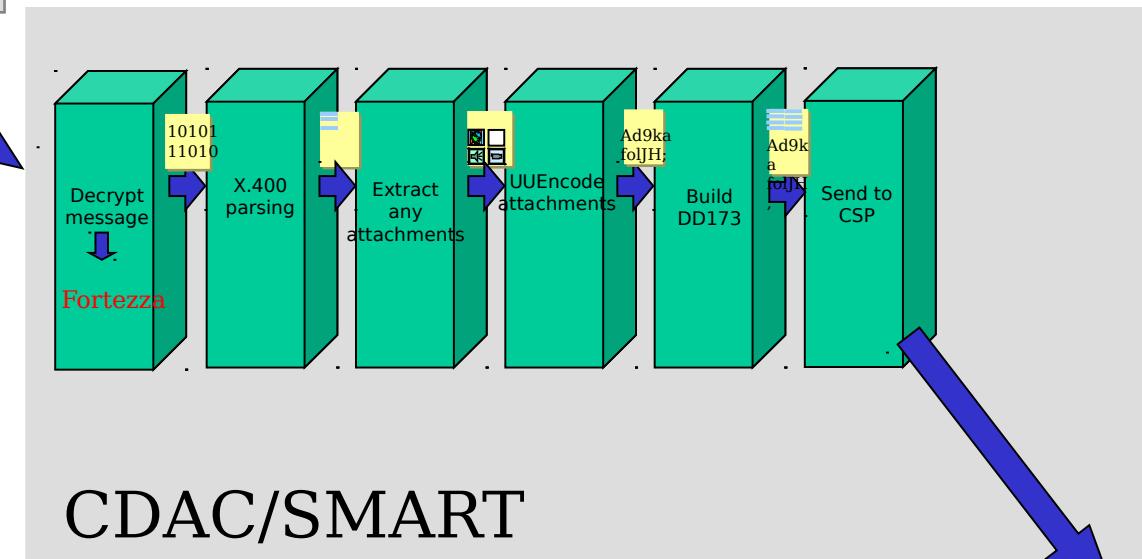
Message Flow - CDAC/SMART

User Agent

DMS



- User creates messages
- Messages may include file enclosures



- DMS passes X.400 messages to CDAC/SMART
- CDAC/SMART converts message to DD173 format
- Messages may include file enclosures
- Enclosures are encoded, attached to the message and passed to CSP

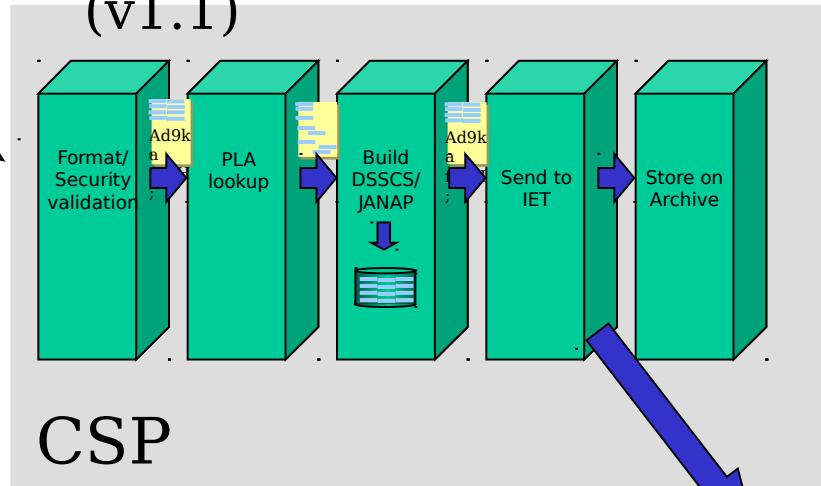


Message Flow - CSP/IET

CDAC/SMART

- CSP passes messages to IET
- Messages may include file enclosures (v1.1)

Ad9k
a
folJH
;



- CSP converts DD173 messages to DSSCS/JANAP format
- Attachments are embedded in the message

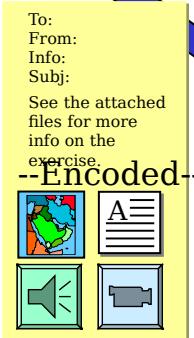
IET



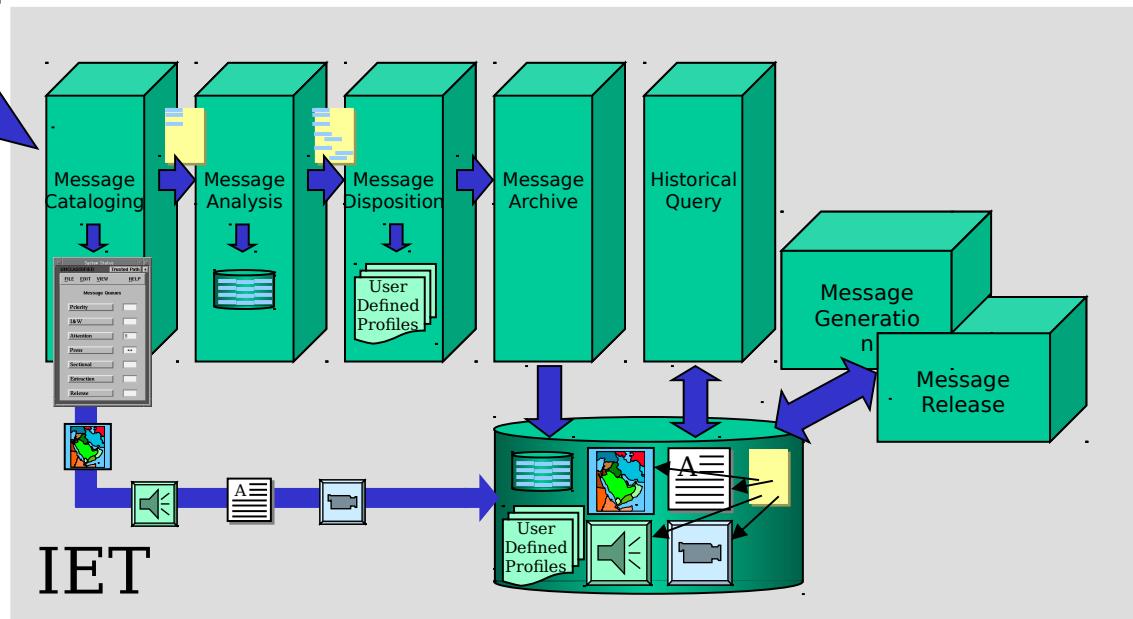
IET Processing

CSP

- CSP passes messages to IET
- Messages may include file enclosures



- Messages are DSSCS/JANAP format
- Enclosures can include text, graphic, imagery, video, binary, audio, etc.



- User queues are updated with new arriving messages
- Message are matched to user profiles
- Messages with enclosures are archived for Historical Query



IET User Interface

The image displays four windows illustrating the IET User Interface:

- Top Left Window:** A "System Status" window titled "UNCLASSIFIED Trusted Path". It shows "Message Queues" with sections for "Priority" and "I&W". Below these are "Attention" and "Press" sections. Buttons include "DISPLAY", "DISPLAY NEXT", "MULTIPLE PICK", and "EXTENDED PICK".
- Middle Left Window:** An "ATTENTION Queue" window titled "UNCLASSIFIED LAW ENFORCEMENT SENSITIVE". It contains a "FILE EDIT VIEW OPTIONS" menu. The main area shows a message header with "Source: STERLING 1//ITD1// DTG : 061605Z DEC 99". The message body includes a working record, a note about another message, and a word document attachment. At the bottom are "Highlight" and "Display" options.
- Top Right Window:** A "StarOffice 5.1" window titled "StarOffice 5.1 - [weather3.jpg]". It displays a map of the United States showing weather conditions. Legend categories include LT RAIN/DRIZZLE, MOD/HVY RAIN, FRZ RAIN/SLEET, LT SNOW/FLUR, MOD/HVY SNOW, and FOG. Specific weather systems are marked with "L" (low pressure) and "H" (high pressure).
- Bottom Right Window:** Another "StarOffice 5.1" window titled "StarOffice 5.1 - [This is a sample Microsoft Word document]". It shows a Microsoft Word document with sample text. The text reads: "This is a sample Microsoft Word document. It can utilize any of the features of word such as different font sizes and characteristics such as bold and underline." The Word ribbon tabs are visible at the bottom.



IET Evolution

- IET v1.0 (Oct 2000 release)
 - Adds Browser interface for profiling, retrospective search and message generation/review
 - Adds BROADSWORD interface for profiling, retrospective search
 - Supports organizational messages
 - Functionally Equivalent to DODIIS AMHS
- IET v1.1 (in development-May 2001 DODIIS testing)
 - Provides capability to profile, store, perform retrospective search of messages with enclosures
 - DII COE Level 5 Compliance/AFDI integration
 - XML support
- IET v1.2 (late FY01)
 - Incorporate advanced message analysis tools (EMCARS)

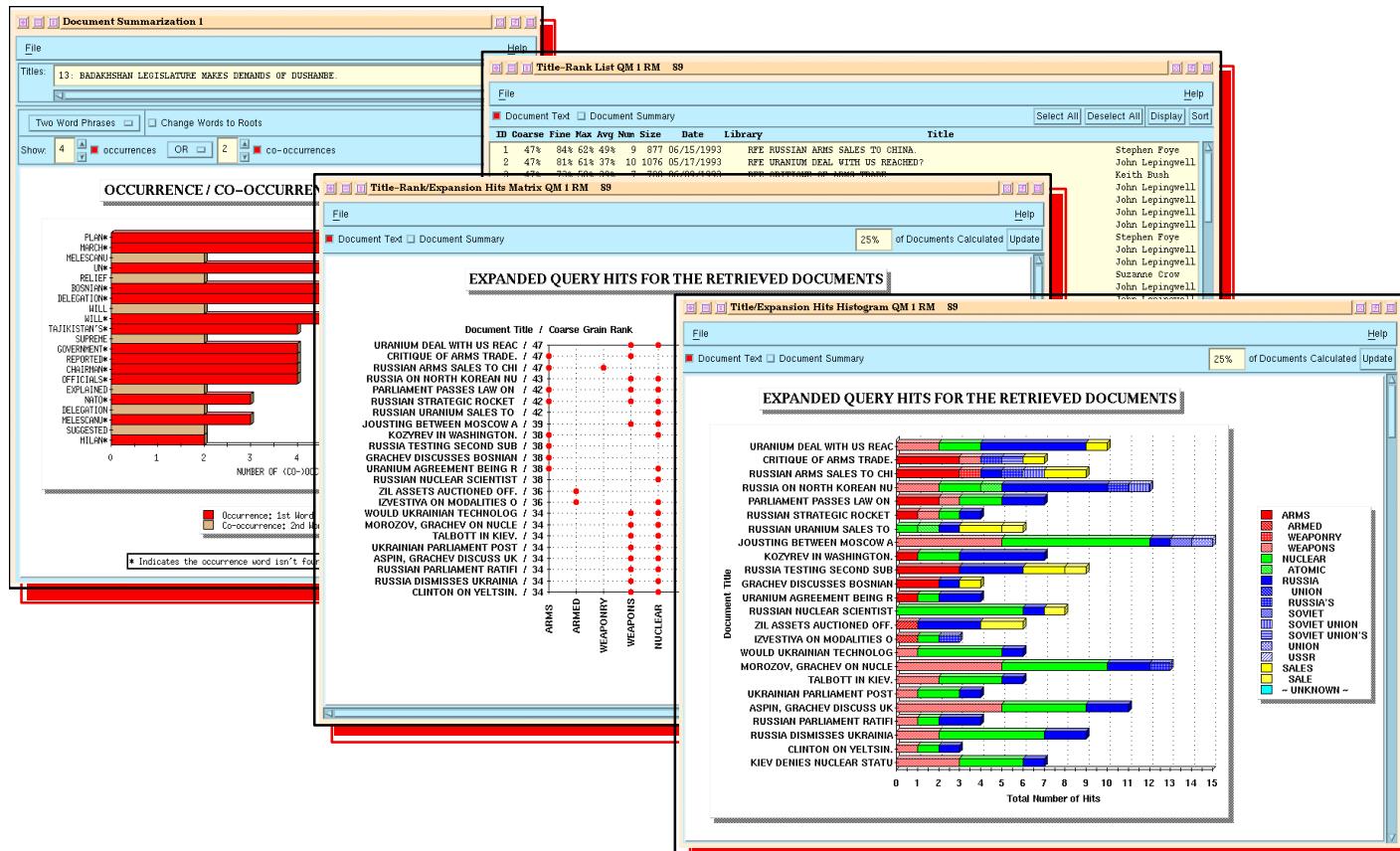
Technology Insertion - Additional Advanced Features (v1.2)



- Automatically handles attachments in over 200 of the most popular commercial proprietary formats
- Visualizations that aid in reducing the time it takes to find the documents that are most relevant to the analyst
- Sophisticated rapid-scan browsing and gisting capabilities for retrieved emails
- Built-in capability to categorize, save, load, edit and re-use both retrospective queries and profiles
- Options to construct and modify existing dictionaries/thesauri into new custom dictionaries/thesauri to enhance recall
- NASA Technical Thesaurus available to facilitate query formulation and provide high precision search (other dictionaries and thesauri are available)



Search Results Visualizations



Visualizations aid in reducing the time it takes to find the documents that are most relevant to the user



Sophisticated Rapid-Scan Browsing and Gisting

The screenshot displays two windows side-by-side. The left window is a "Single Document Text Viewer QM 4 RM 1" showing an email message with subject "Re: [NTSEC] Transplant passwords from UNIX to NT" and from "Christine Keogh <Christine.Keogh@anu.edu.au>". It includes fields for FILE_NAME, Doc_Type, Num_Attachments, DOC_LENGTH, and From_List. Below the message content, there are buttons for "Previous" (99), "Next" (112), and "Goto". The right window is the "ECT (RW6.8.2) 1.0.3 Email Culling Tool" with a toolbar for Run DCARS, Path Selection, Processing Stages, and Previous Results. It shows a list of queries: 1. Computer terms (Deep), 2. afmil, 3. commerce query, 4. communication words, and 5. email terms. A "Clear Console" button is available, and the stage is shown as "Querying Completed". The console output shows log messages like "Cleaning Exceptions Directory: Completed" and a list of 16 files to process, including file1.set2.SW3 through file2j.set2.SW3.

```
#####
# Cleaning Exceptions Directory: Completed #
# Tue Feb 8 11:02:48 #
#####
BEGIN: /vob_storage/avail_vobs/dcars/rw6.x/ect/util/Remainder

Total number of emails to process:
16 : /home/paella/EMCARs_testing/input/
/vob_storage/avail_vobs/dcars/rw6.x/ect/util/bin/h_and_l

16) file1.set2.SW3
15) file1inc.set2.SW3
14) file1j.set2.SW3
13) file2e.set2.SW3
12) file2inc.set2.SW3
11) file2j.set2.SW3
```

Sophisticated rapid-scan browsing and gisting capabilities
for retrieved emails enhances rapid drill down



Deployment Options/Costs

- New standalone server (parts list available)
 - Small (\$20k) - 513MIB (Fort Gordon)
 - Medium (\$31k) - TRANSCOM, EUCOM, SPACECOM, 480IG
 - Large (\$67k) - SOUTHCOM, SOCOM, CENTCOM, JAC
- Reuse of existing Sun hardware (per discussions with individual sites)
- Installation cost approximately \$28k/per site
 - Includes software installation/configuration, data migration from legacy system (AMHS/DMFE), security accreditation testing, training
- No annual user fee for IET software maintenance/upgrades per 497IOG/IND



IET Activities

- IET v1.0 Beta II Testing
 - Successfully completed JITF testing, Aug 2000
 - Successfully completed Beta I testing at USSTRATCOM 5-7 Sep 2000
 - Conducting site demonstrations per 497IOG direction
- 497IOG Site Install, Oct 2000
- IET v1.1 Requirements Review, 497IOG, Jan 2001
- AFRL developing roadmap for information extraction technology insertion
 - Email Content and Retrieval System (EMCARS)
 - Generic Intelligence Processor (GIP)



Summary

- Information Extraction Tool provides DMS-era message handling functionality
 - Archiving
 - Profiling
 - Retrospective search
 - Message generation
 - Accommodates DMS messages with enclosures
- IET v1.0 (DMFE v7.0) ready for deployment Oct 2000
- 497IOG has recommended IET as replacement for DODIIS AMHS



AFRL/IFEB POCs

- Chief, IDHS Program Office - John Vona, 315-330-3601 (DSN587), vonaj@rl.af.mil
- Program Manager - John Baumler, 315-330-3979 (DSN587), baumlerj@rl.af.mil
- Lead Engineer - Mike Corelli, 315-330-7525 (DSN587), corellim@rl.af.mil
- AFRL/IFEB home page
 - Internet - <http://www.if.afrl.af.mil/programs/iet/>